DOCUMENT RESUME

ED 279 150 FL 016 173

AUTHOR Enkvist, Nils Erik; von Wright, Marianne
TITLE Problems in the Study of Textual Factors in Topicalization.
PUB DATE 78
NOTE 29p.; In: Andersson, Erik, Ed. Working Papers on Computer Processing of Syntactic Data. Abo, Research Institute of the Abo Akademi Foundation, 1978. p45-71.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Coherence; Computational Linguistics; Data Processing; Deep Structure; English; *Grammatical Acceptability; Language Patterns; *Language Styles; Paragraphs; *Research Methodology; *Sentence Structure; *Structural Analysis (Linguistics); Surface Structure; Syntax
IDENTIFIERS *Markedness; *Thematic Analysis

ABSTRACT

Certain word-order patterns are more basic and less marked than others. The more strongly marked a pattern seems to be in isolation, the stronger must be the contextual forces motivating its use, if it is to seem natural in a text. Various topicalizations (of adverbials, objects, and parts of verb phrases, for example) need various degrees of motivation. If the motives are strong enough, they may justify the textual fit even of strongly marked sentence patterns. One characteristic of poetic license is the freedom to use strongly marked thematic patterns. Different languages, even relatively closely related ones, give very different degrees of importance to the informational organization of the sentence. In foreign language teaching, differences in the handling of information flow in the course and target languages deserve more explicit attention, especially where the languages use different mechanisms. In translation, more information about the relative markedness of thematically different sentence patterns and about the forces improving or impairing textual fit would be helpful. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279150

Nils Erik Enkvist
Marianne von Wright
Abo Akademi

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
N. Enkvist
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PROBLEMS IN THE STUDY OF TEXTUAL FACTORS IN TOPICALIZATION*

1

One of the conspicuous developments in linguistics over the past few years has been a sharpening of the tools used to describe 'grammaticality', 'acceptability', and 'appropriateness'. If our reading of the recent history of linguistics is right, this owes to two needs. First, when studying corpus materials, linguists have had to learn to deal with actual parole or performance, and with all the deviance inherent in impromptu communication and in texts using deviant patterns for special effects, such as advertisements or modern poems. They have had to decide what was normal enough to be described in the grammar, and what was marginal or deviant. Secondly, when trying to set up generative models, grammarians have been compelled to evaluate the output of their rules to see whether the rules were good or bad. Consequently, the concepts and terminology have gained in precision. In one commonly used sense, 'grammaticality' indicates the degree of correspondence between a certain expression and the rules of a certain grammar. 'Acceptability' on the other hand is measured through elicitation experiments, perhaps simply on a graded scale involving judgements such as 'acceptable', 'question-

* Marianne von Wright has written Section 5 on topicalizations, using data drawn from her unpublished M.A. thesis (Some Textual Factors in Topicalization. A Preliminary Corpus Study, Abo Akademi, 1977). The remainder of the paper emanates from Nils Erik Enkvist, who also read a preliminary version of it for the Systemic Grammar Workshop at the Hatfield Polytechnic in July, 1977. We are grateful to Erik Andersson for constructive discussions and friendly editing.

BEST COPY AVAILABLE

FL 016 143

able' and 'unacceptable', perhaps in terms of operations where the informants are expected to manipulate sentences. 'Appropriateness' links an expression to a situation: an expression may be appropriate in a barrack-room but shocking or startling in a speech from the throne, and vice versa.

In recent linguistics it has also become increasingly clear that sentences are not autonomous. Sentences are part of texts: the one-sentence text is an instance of an extreme textual minimum. And in real-life communication sentences occur as parts of discourses. Students of language have further learned something, and are in the process of learning more, about those textual and discoursal forces that affect the form of the individual sentence. Thus the job of a sentence is to contribute to the flow of information that runs through the text. In a well-formed text, sentences have to connect both with what has gone before and with what comes after. This makes text-initial sentences different from text-medial ones, and perhaps also section-initial sentences different from section-medial sentences. Further, sentences should make an optimal distinction between old information -- that is, information presumably known to the receiver of the message either because of culture, experience, the immediate environment of the speech act, or previous mention in the discourse -- and new information. If the text is to qualify as stylistically homogeneous, and in that sense stylistically well formed, its individual sentences must be formed in stylistic harmony with their textual environment. They must be stylistically appropriate. And at the higher level of text organization and narratological form, sentences may simultaneously have to obey certain higher structural principles. In metrically regular poetry, for instance, sentences must be shaped to fit the metre; the final couplet of a Shakespearean sonnet should be something like an epigrammatic summary of a sentiment or conceit; and the beginning of a business letter should conform to specific patterns that are taught to prospective secretaries and correspondents.

Though we have begun by recalling the range of textual and discoursal forces that may affect the acceptability of a sentence in context, our purpose is not to speak about them all in one brief paper. Our aim is merely to remind the reader of the fact that uncontextualized acceptability ratings answer the question "Is expression A ever acceptable?" rather than "Is expression A acceptable in context B?". But contextualized acceptabilities are obviously very important too.

In fact one might speculate and hazard the view that all acceptability ratings may ultimately be based on tacit contextualization. If one is asked, "Is expression A ever acceptable?", a plausible strategy would be to try to think of a context in which A would be all right. If one can conjure forth such a context, A becomes acceptable; but if one cannot imagine a context in which A is all right, the expression must fail the test.

Apart from such a reminder of the importance of contextualization for acceptability, we shall not go into contextual restrictions of cross-reference and coreference, into problems of stylistic homogeneity, and of narratological structure or poetic form. We shall not for instance look at the types of restrictions that obtain for text-initial sentences, the shifts in reference -- for instance in genericity -- that destroy coreferentiality between formally related or identical lexemes, and the like. We shall instead outline some principles related to the ways in which thematic organization (theme-rheme or topic-comment structure, functional sentence perspective) contributes to the contextual acceptability of a sentence.

2

One type of textual adjustment, then, is the choice between thematic variants. This statement assumes that we are capable of distinguishing thematic variants (such as (1) and (2) below) from other rearrangements of sentences: thematic variants must be cognitively equivalent but may differ in their presuppositions as to old and new information. Then, intuition and considerable empirical evidence suggest that some thematic variants strike us as being more marked than others when presented in isolation. And the point we wish to make is that the justification for the use of the more strongly marked thematic variants is to be sought in situational and textual parameters. The factors affecting choice of thematic variant reside not in the sentence itself but in its environment.

For instance, when given in isolation, the following sentences out of Wyndham Lewis's The Revenge for Love (Penguin ed., 1972) strike native speakers of English as odd, needlessly elaborate, rare, and artificial. If they were found in a paper written by somebody learning English, many

teachers would be likely to correct them and give them a less marked form:

(1) (a) Be touched by this woman he must on occasion. (P. 55.)
(b) Interfere he was unable to. (P. 372.)
(c) Two distinct and watertight Margots, Victor considered he could manage. (P. 323.)

Such less marked forms would be

(2) (a) He must on occasion be touched by this woman.
(b) He was unable to interfere.
(c) Victor considered he could manage two distinct and watertight Margots.

In the discussion of such examples we shall need two conceptual tools, both of which were already mentioned. One such concept is markedness, the other is thematic variation. An explanation of how these terms will be used here is therefore in order.

3

Markedness is a term which has been used in at least four different senses. The first, and oldest and most fundamental, sense is that the marked member of an opposition is that member which has a marking feature absent from the other, unmarked, member. Such marking features may be phonological (in the opposition p/b, b has voice which p lacks), morphological (in kick/kicked, the past and participle are marked with -ed), syntactic (in the opposition active/passive, the passive carries the markers be + participle ending), lexical and semantic (in lion/lioness, lioness has the lexical marker -ess and the semantic marker [+feminine]). Note, however, that the choice of marking feature is subject to the linguist's manipulation. In the opposition p/b, we can turn the p into the marked member if we describe the contrast with the aid of markers such as [±fortis].

The second sense in which markedness might be defined is in terms of distribution patterns: wide distribution suggests less, narrow distribution more markedness. Of the two English plural morphemes -s and -n, -s can be defined as the unmarked one because its distribution is wider than that of the -n (as in oxen, children etc.). The same applies to features at other levels: in conclusions of letters for instance, yours sincerely is more widely applicable than yours faithfully, and therefore

less marked in this particular sense. The third sense of markedness is related to the second sense and involves actual frequency counts or frequency estimates: high frequency suggests low degrees of markedness.

Finally, markedness has sometimes been regarded as an intuitive concept. The least marked structure then becomes the one that a suitably selected group of informants -- and even the analyst alone -- would characterize as the most common, ordinary, usual, inconspicuous, or fluent variant. In this sense, markedness becomes related to what Russian formalists and Prague-School critics call 'foregrounding': a marked structure calls attention to itself, whereas an unmarked structure has, metaphorically speaking, a kind of mimicry that helps to blend it into its background.

All these four types of criteria measure relative degrees of markedness in what might be called a paradigmatic dimension. But we might also conceive of a context-sensitive, syntagmatic measure of markedness. Thus in the sequence VCV, where V stands for a voiced sound and C for a consonant, we might regard voiced consonants as less marked than unvoiced ones: a sequence such as apa would then be more marked than a sequence such as aba. In Firthian terms, frequent changes of prosodies suggest more markedness than less frequent changes.

Indeed for the student of thematic markedness in context this distinction between paradigmatic, uncontextualized markedness and syntagmatic, contextualized markedness is very important. Thus if we present informants with the two sentences

(3) (a) On the bald streets breaks the blank day.
(b) The blank day breaks on the bald streets.

and ask him to give them relative ranks for markedness, many informants will regard (3b) with the order SVAdv as the more common and thus less marked variant. But in context, the criteria may change. What Tennyson actually wrote in In Memoriam was

VII. Dark house, by which once more I stand
Here in the long unlovely street,
Doors, where my heart was used to beat
So quickly, waiting for a hand,

A hand that can be clasp'd no more -
Behold me, for I cannot sleep,
And like a guilty thing I creep
At earliest morning to the door.

He is not here; but far away
  The noise of life begins again,
  And ghastly thro' the drizzling rain
On the bald streets breaks the blank day.

In its proper context, (3a) seems more natural than (3b), and for several reasons. Rhyme is one of them. Then, the street scenery is old information (referring to the Hallams' residence in Wimpole Street). Also the streets form a backdrop for the entering dawn and have a temporal priority over the blank day. Thus, in context, what in paradigmatic isolation was the syntactically more marked variant actually becomes the textually less marked variant.

This is of course only another way of saying that the justification for the use of syntactically marked structures is to be found in context: the forces justifying the syntactic marking are contextual. All the same we may see a dualism between syntactic, paradigmatic markedness and contextual, syntagmatic markedness. With an animistic metaphor we might say that the syntax of a sentence contains a certain inertia resisting increases in markedness, and that contextual forces are needed to overcome this resistance. As there seems to be such a dualism, the different and perhaps opposing forces deserve labels of their own. We shall therefore speak of MARKEDNESS for short, or SYNTACTIC MARKEDNESS for clarity, to indicate uncontextualized markedness. And we shall use the term TEXTUAL FIT to indicate contextual adjustment, often of a kind which increases syntactic markedness. In (3a) for instance the textual fit has been improved by increasing syntactic markedness.

4

Our next problem is how to desribe the syntactic patterns whose thematic markedness we wish to measure. To begin with, we have recently begun to realize that thematic marking and information dynamics play very different roles in different languages: the contrast between theme-rheme-dominated and subject-predicate-dominated languages may even serve as the basis of a language typology (Li and Thompson 1976). In English, there are different methods of organizing the information flow within a sentence. The purpose of this organization is, roughly speaking, to make a distinction between old and new information, and in an even wider perspective to facilitate the exchange of information by eliminating alternatives in a

certain specific order. As information theorists tell us, information consists of the elimination of alternatives and in the specification of certain possible worlds as contrasted with other worlds. And the thematic organization of the sentence regulates the order in which alternatives are eliminated. In English, grammars have discussed related matters under labels such as word order, functional sentence perspective, theme and rheme or topic and comment, focus, weight, and the like.

Now a reminder of a few well-known facts. To organize the information pattern of a sentence and to adjust textual fit, English has a number of devices to offer. Compare

(4) (a) Peter bought the car from John.
(b) John sold the car to Peter.

(5) (a) The ashtray is on top of the book.
(b) The book is under the ashtray.
(c) *The bed is under the ashtray.

(6) (a) John was driving the car.
(b) The car was being driven by John.

(7) (a) I had not seen him for a long time.
(b) Him I had not seen for a long time.

(8) (a) JOHN I had not seen for a long time.
(b) It was John that I had not seen for a long time.

In (4) and (5), the thematic adjustments have been carried out with the aid of lexemes through a choice between 'converses'. The choice of the appropriate lexical item makes it possible to place a certain element (theme) into subject position and thus into one of the initial slots in the linear string. (5c) shows us an instance of a restriction (and presumably a pragmatic rather than syntactic one) on such conversion. In the active construction in (6a), the car is the surface subject. The passive is thus an example of a structure used to change the thematic organization of a sentence through a change in syntactic structure. In (7), the difference between (a) and (b) is not one in the syntactic function: I is the subject and him the object in both. The difference is one of word order, or rather of linearization (because what is usually called 'word order' affects not only words but also word groups and clauses): in (7b), him occurs in front position, and many grammarians would say that it has been fronted. In (8), John has been marked, in (a) by stress and intonation (here indicated with capitals) and in (b) through clefting. Such instances of marked focus usually indicate that the focally-marked item contains new information though it occurs in a place where we normally expect old

information. There is an interplay between the assignment of syntactic functions, the fronting or postponement of items, and the possibilities of marking focus through stress-cum-intonation or through cleft. But this is a question we shall not pursue here.

Apart from instances with marked focus such as those in (8), we can, then, classify thematic operations into three major classes, each of which deserves a label. But when labelling phenomena of this kind, we are on dangerous ground, because terms such as theme and rheme, topic and comment, and focus and emphasis have been used in notoriously different senses. To many scholars, topic and theme are synonyms, and so are comment and rheme. Theme and rheme are favourite terms in Prague and Britain, topic and comment in the U.S., with many individual exceptions both ways. But there is no unanimity in the definition of such terms. To some linguists, the difference is a matter of deep semantics or underlying logic. Themes might thus be defined as the elements with maximum scope in the sentence. To others, thematic adjustments are cosmetic operations close to the surface. Some linguists equate theme or topic with old, and rheme or comment with new information; Jan Firbas speaks about a scale of Communicative Dynamism, themes becoming minima and rhemes maxima of CD, with a possibility of gradations between the extremes. Yet others prefer more formal syntactic definitions, defining themes as initial and rhemes as final elements. In view of the multiplicity of such definitions, all students of information structure should be careful to reexamine their own use of these terms. In fact there is little sense in arguing for one definition and against another out of context. The choice of definition may very much depend on the aspect of linguistic structure one is especially interested in. Those working with semantics in a logical framework will thus see these matters in a perspective different from that of a student of literary style. All definitions may be justified within their own frame of reference. They may even be brought into some kind of harmony.

There are good reasons for departing from the tradition that regards topic and theme as synonyms, and comment and rheme as synonyms. We shall here define thematizations as those thematizing and rhematizing operations that affect the assignment of surface-syntactic functions, and topicalizations as those topicalizing and commentizing operations that merely move items about in the sentence, without affecting their syntactic functions. In (4) and (5), we can thus speak about lexical thematizations, about thematizations achieved through the choice of lexemes. In

(6), we can speak about syntactic thematization, that is, the change of thematic roles through a syntactic operation (here, passivization) which changes the syntactic roles of constituents (object into subject, subject into agent). And in (7), we have an instance of a topicalization, a fronting of him without further syntactic change. In this terminology, then, every sentence (or at least most sentences) have a theme and a rheme. But only those sentences in which an item has been moved -- fronted or postponed -- from its basic, unmarked position have topics or comments. A topic is thus part of the theme, a comment part of the rheme. Depending on the motive underlying these operations we may further try to distinguish between thematizations whose purpose is to thematize a given item, and rhematizations whose purpose is to rhematize a given item; and similarly between topicalizations and commentizations. But such distinctions presuppose an analysis of textual parameters. Also the number of degrees of freedom may confuse the issue: in the Swedish example

(9) (a) Kalle läste boken.[1] ('Charlie read the book.')
(b) Boken läste Kalle. ('The book read Charlie.')

the choice of the SVO pattern can depend both on a desire to thematize Kalle and rhematize boken, and the choice of the OVS pattern similarly on a simultaneous desire to topicalize the object and commentize the subject. But once the theme or topic has been chosen, there is no freedom in the choice of rheme or comment: the assignment of rheme or comment follows automatically. Indeed both motives, thematization and rhematization, may conspire to work together: it would be wrong to ask whether the structure of a sentence is due exclusively to one or the other, when both forces act simultaneously towards the same result.

Another possible objection to this distinction between thematizing operations which affect syntactic functions and topicalizing operations which do not is that the term 'syntactic function' is a bit slippery. What about the existential construction for instance? The reason why people use structures such as

(10) There were two books on the table.

---

1 As läsa, 'to read', requires human subjects (except in metaphors), the OVS pattern is acceptable and unambiguous in Swedish. In general, Swedish allows object fronting whenever there is no risk of confusing a fronted object with a subject. For examples, see Sköld (1970).

is thematic insofar as its purpose is to avoid placing new information (two books) into sentence-initial position. But two books is presumably still a subject, there being semantically empty. The same problem can be posed about extrapositions of the kinds illustrated by

(11) It is well known that Columbus discovered America.
(12) It is difficult to please Sheila.

Should such structures be called thematizations or topicalizations, or should they be given a separate category of their own, where semantically empty morphemes (such as anticipatory there and it) fill theme slots to enable semantically loaded elements to go into the rheme? The answer will depend on the criteria one wishes to adopt for deciding between these alternative solutions. The distinction between thematizations and topicalizations needs further study on points such as these.

Within this general framework one can now go on to describe and classify thematizing and topicalizing operations with the aid of the grammar model of one's choice. One must then decide whether to describe word-order patterns in relative terms (subject before verb, verb before subject etc.), in terms of slots and fillers, in terms of 'fields' such as Diderichsen's, or with transformational models, which should ideally generate sentences with basic unmarked word-order patterns and then provide transformations through which more marked structures are brought about. In this light it is instructive to read different grammars, not least recent British ones and transformational ones, and see which of their operations bear on thematic ordering. The grammars of Halliday and his followers as well as the Grammar of Contemporary English by Randolph Quirk and his collaborators are in fact pioneers in paying explicit attention to thematic matters. American transformationalists, Emonds for instance to mention one concrete example, list a number of thematically relevant operations without expressly relating them to a special thematic subsystem of language. It is up to the reader to cull thematic operations out of such books and to decide on their thematic relevance.

5

We can try both of the linguist's usual two ways of collecting actual data to see how this background machinery can be put to empirical

use. One way is to analyze a corpus. The other is to conduct elicitation experiments, either simply by using one's ubiquitous informant -- oneself -- or more elaborately by interviewing groups of suitably selected people. As we have been involved in some experimentation with both, a few preliminary remarks about methods and results might be of interest. It should be made quite clear, however, that what we have done has been very much in the nature of pilot experimentation, and the purpose has been to see how experiments work out rather than to collect data on which to build firm results.

The corpus study was based on thematically marked sentences from Wyndham Lewis's novel The Revenge for Love (abbreviated WL in the examples below), Angus Wilson's novel The Middle Age of Mrs Eliot, British and American newspapers, journals and other casual reading. The resulting corpus contained 411 sentences, out of which Wyndham Lewis provided 225 sentences, i.e. more than half of the entire corpus. In addition to topicalizations, the corpus included clefts, pseudo-clefts, extrapositions, and dislocations. The stylistic representativity of the corpus was wholly uncontrolled. Our experiments suggest, however, that the density of certain types of topicalizations is a style marker. Wyndham Lewis for one seems to be characteristically fond of topicalization, and so are some journalists and contributors to highbrow journals. In contrast, a sample of slightly-edited transcripts of spoken dialogue from The American Issues Forum, a record of transcribed, academic discussions, proved to contain surprisingly few topicalizations, whereas dislocations, extrapositions and pseudo-clefts occurred more freely in these transcripts.

The sentences of the corpus were then coded for computer treatment. The information coded was of the following kind (with the numbers referring to columns on the punch cards):

(1-3) Identification of sentence in corpus, (4) position of the clause containing the topicalization in the sentence, (5) type of clause, (6) sentence type (active transitive, active intransitive, existential, agentful passive, agentless passive), (7) length of fronted constituent, (8) type of fronted constituent, (9) givenness of fronted constituent (just mentioned, mentioned further back, be/have relation, indexal implication, presupposed as known, generic, pseudogiven, new, uncertain, irrelevant), (10) length of postponed element, (11) type of postponed element, (12) givenness of postponed element, (13) length of final element, (14) syntactic function of topicalized constituent, (15) trigger of topicalization (direct connection, emphasis, contrast or correction, indirect because of postponement of heavy subject, object, or comple-

ment, weak direct and indirect, iconic, disambiguation), (16) subject postponement (it/there insertion, simple subject postponement, extraposition of subject clause, extraposition with split subject clause, (17) verb shift, (18) cleft, (19) dislocation, (20) raising, (21) total number of movement transformations.

The terms of the coding key are derived from concepts used in general linguistic theory; still some of the criteria used may need more detailed explanations (cf. also Enkvist 1975).

Columns (9) and (12) rate the information content of topicalized and postponed, or commentized, constituents. A 'just mentioned' element is coreferent or cross-referent with an item which has occurred earlier in the same section of the text. In a newspaper or periodical such a section consisted of a paragraph; in a novel, of a sequence of sentences describing a single scene or a coherent series of events. 'Mentioned further back' refers to a preceding section. The adverbial of place, at the window, in (13) stands in a 'be/have relation' to the taxi:

(13) The taxi arrived, and at the window was the face of Jack ... (WL p. 221.)

Hyperonymic-hyponymic relations, e.g. flower - tulip, were also included under this heading.

'Indexal implication' is the kind of relationship that holds for instance between sun and shade, finger and ring. 'New' information can be either a new discourse referent or an item which is new in that particular context. In the present study, the value 'irrelevant' was assigned to items which were difficult to place on a scale from given to new, for instance verbs and adverbs of manner and degree. A closer examination of the context often revealed links from a topicalized verb to a preceding sentence, although these links were frequently difficult to categorize according to the present coding key. An obvious link, which could have been classified as a 'be/have relation', was found between be touched by this woman and contact with Sister Teresa in (14):

(14) And again contact with Sister Teresa was imperative - which made him more sick than ever. Be touched by this woman he must on occasion, that ... (WL p. 55.)

But as mentioned above, the givenness of the fronted part of a verb phrase was classified as irrelevant. An adverb of manner, on the other hand, for

instance objectively in the last sentence of the passage in (15) presents more of a problem.

(15) Percy did not move. Half his face, shining with sweat, was bathed in the moonlight. /.../ There was an effort on his part to remember his instructions, in the event of anything occurring. He was to be sick. /.../ He was sleepy, however. He was in quite unexpected harmony, all of a sudden, with Nature - /.../ Don Alvaro shook his head reproachfully.
'I am offended, Don Percy!'
Objectively Don Percy considered this apparition, ... (WL pp. 49-50)

The semantic connection from the adverb to the context is purely intuitive and not describable with the terms used earlier. Also, as many adverbials, especially adverbs of manner and frequency, may fairly freely take either front, medial or end position (Quirk et al. 1972), the term 'irrelevant' seemed justifiable for adverbials that would not fit into any of the categories of givenness.

Within column (14) adverbials were classified as (a) valency adverbials, (b) setting adverbials, (c) other continuous adverbials or (d) separated elements, on the following principles.

(a) Valency adverbials are closely connected with the verb and consequently deeply embedded in the hierarchical structure. In traditional terms they can be classified as adverbials of time, place or manner which are part of the semantic specification of the verb. They can be important for the acceptability of a sentence; if a tightly-bound valency adverbial is removed the sentence becomes unacceptable or changes meaning. Cf. on the table and by car in (16) and (17):

(16) Put the book on the table.
*Put the book.

(17) The boy goes to school by car.
?The boy goes to school.
?The boy goes.

Valency adverbials specify the action itself and depict relations and views that are essential for that particular utterance (Enkvist 1976). In this respect they differ from (b) setting adverbials, which form the temporal and spatial frame for the action, and as such belong to the whole clause, as in (18):

(18) After a week he was still there in the company of a half-dozen insignificant criminals and vagrants. (WL p.375.)

Classifying adverbials into two categories, valency and setting, is in fact an oversimplification: a more delicate classification would have to reckon with a number of gradations on a scale indicating the boundness of an adverbial to its verb.

Along another dimension, adverbials can be divided into two main groups, continuous ones which are integrated into the clause and form part of the same tone group, and separated ones which form a tone group of their own divided from the rest of the clause by an intonation break in spoken language and by a comma in writing. In the present study, all continuous adverbials that were not valency or setting adverbials were classified as (c) other continuous adverbials, while separated adverbials formed the fourth adverbial category.

'Direct connection' as a possible trigger of topicalization in column 13 refers to a topicalization where a close link exists between the fronted constituent either in the form of co-reference or some other kind of close textual cohesion. 'Emphasis' was used in sentences where an important constituent, either given or new, was picked out for special emphasis, and 'contrast or correction' when the topicalized constituent formed a contrast to or a correction of a previous one. In sentences where we felt that the rhematic part of the sentence was needed for important (new) information, the trigger was classified as indirect, as in (19) below; if, on the other hand, an element had been placed in initial position to avoid a clustering of constituents after the verb phrase, this was called a 'direct and indirect' reason for the topicalization, as in (20).

(19) At the farther extremity of this reception hall was a door which led out into the street. (WL p. 14.)

(20) At the Pole she would be at home with Victor, as ... (WL p. 349.)

'Iconic' trigger meant that between two sentences there existed an iconic link, i.e., a constituent or a clause acted as a picture of an adjacent constituent or clause (Enkvist 1975). In (21 a-b), both sentences begin with <u>here</u> and have the structure adverbial + <u>be</u> + subject.

(21) (a) Here was the bearer of offensively large bouquets to ladies.
(b) Here was the celebrated Red beau, back from the battle ... (WL p. 217.)

If the topicalized element had been placed initially to avoid an ambiguous sentence, the trigger mechanism was coded as 'disambiguation':

(22) With a bloodshot eye he was following the flight of the king to Varennes. (WL p. 57.)

However, it should be stressed that the borderlines between the categories are vague, and that the coding often provided difficulties and sometimes led to unreliable results. One of our experiences was that we strained too hard to assign patterns to single causes, without allowing sufficiently for conspiracies involving several simultaneous causes.

With alpha-numeric coding, all this information could be fitted into 21 columns on a punch card. We soon noticed the difficulties in setting up a consistent coding policy. Inevitably, similar structures got coded differently as one's views changed during the progress of the work. Also there were structures not foreseen in advance, which necessitated revisions in the coding key. One difficulty consisted of assigning givenness ratings to adverbs. Another concerned the reasons for topicalizations. It was, for instance, often problematic to decide whether the link to the preceding context was one of plain connection without any special marking of emphasis or contrastiveness. Clues for the decisions had to be found in the text. Such clues were to a great extent intuitive and could involve assessments of the tone of a particular section of text, or decisions as to whether an element with which the topicalized element might be contrasted had been mentioned not too far back. Similarly, indirect reasons and combined direct and indirect ones tended to overlap. Disambiguations were often difficult to spot and some instances may have been overlooked. Since many topicalizations could be classified under several of the trigger types, a more explicit hierarchy between the triggers should have been established. Iconic links, for instance, always involve a second reason for topicalization; compare sentences (21a-b) above, where here also serves as a direct (emphatic?) connection to the previous context. Instances with iconic links were, however, only coded as such.

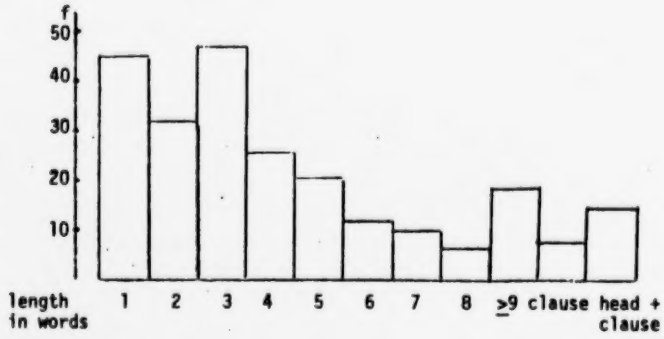
Using a cross-tabulating programme developed for studies of syntax in the Text Linguistics Research Group (Kohonen and Salmela, this volume), a number of cross-tabulations were produced by computer. The corpus was uncontrolled as to sample representativity and limited to 411 sentences. The coding key was restricted to 21 variables, and no figures were available as to the number of untopicalized sentences or sentence

elements in the texts. Therefore the cross-tabulations could only be used as rough indicators of relative frequencies and as starting-points for rechecks of the original sentences and their textual environments. One valuable experience was that once the programme had been developed, the technical operations of punching and computer analysis of a corpus this size proved fast and inexpensive. The corpus can therefore easily be extended. However, all that went into the computer was the alpha-numeric coding of the syntactic patterns, which had been analysed manually. The text proper was thus handled manually, which of course saved a great deal of punching and computer time.

The results of the cross-tabulations were used to formulate tentative answers to the questions: what type of constituent is placed in topic as against comment position, and why is a constituent moved from an unmarked to a more marked position? We shall here give some examples of the type of information which can be extracted from a study of this type.

Diagram 1, frequencies of topics of different length, below, is based on 243 topicalized sentences and shows a bias for short fronted elements consisting of 1-3 words. However, it must be borne in mind that we did not produce the corresponding figures for all untopicalized elements, which could have been used as a basis for comparison. Short elements might be equally frequent in other positions. The short fronted

Diagram 1. Frequencies of topics of different length (N=243).



elements are typically adverbs and anaphoric pronouns, e.g. (23) and (24), and prepositional phrases, e.g. (25) and (26):

(23) Always he had respected the English from afar. (WL p.12.)

(24) That it was important to get fixed in one's head - ... (WL p. 12.)

(25) With eagerness he boiled himself a kettle ... (WL p.83.)

(26) Among the heavyweights there were so few who could be trusted even to understudy the 'giants'. (WL p.254.)

If Diagram 1 is compared with Diagram 2, below, which gives the distribution of final elements of different lengths, we notice that the overall profile of the two diagrams is very much alike. There is a tendency to reserve end position more frequently for long constituents, however.

Diagram 2. Frequencies of final elements of different length (N=259).

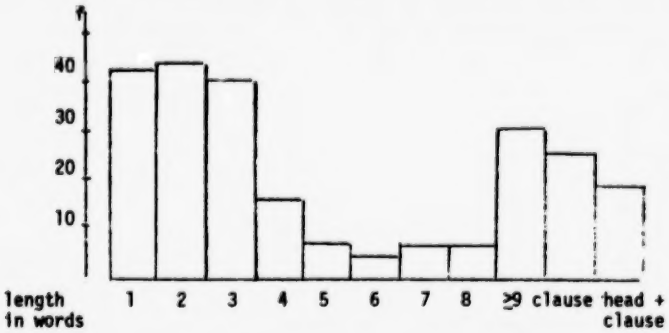


Diagram 3, next page, distribution of the givenness of the topicalized constituent, supports the hypothesis that given information is placed at the beginning of the sentence. In order to get a clearer picture of the distribution of givenness within the corpus, the 51 instances of uncertain information content were omitted from this diagram. They are, however, included in Diagram 4, next page, which shows the frequencies of all the coded types of givenness. The high proportion of topics containing new information or linked to the preceding context through a be/have relation should be noticed. It may, however, be partly due to Wyndham Lewis's eccentricities, partly to doubtful coding, too restricted contexts, and the like.

Diagram 3. Distribution of given and new information in topicalized constituents (N=243).

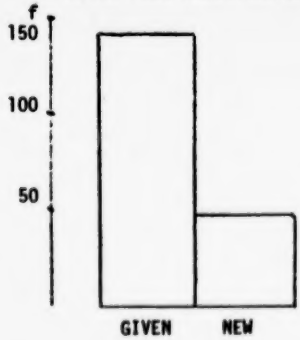
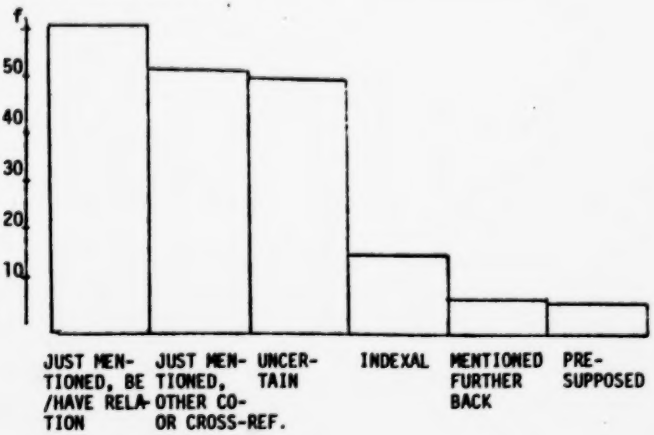


Diagram 4. Distribution of givenness type for topicalized constituents (N=243).



The corpus contained 53 sentences where one constituent could be defined as having been moved into topic and another one into comment position. In more than half of these instances, 32 sentences, the postponed element contained new information, as shown in Diagrams 5 and 6, next page. This figure might be considered surprisingly low, as the most prominent intuitive reason for a postponement is the desire to place new information late in the sentence. The postponed constituents containing

Diagram 5. Distribution of given and new information in postponed elements (N=53).

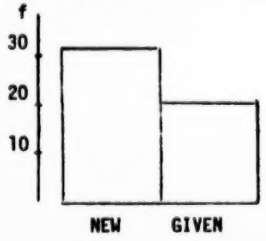
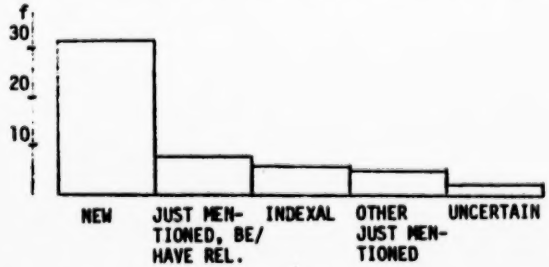


Diagram 6. Distribution of givenness type of postponed element (N=53).



different types of given information might, on the other hand, have followed the principles of end focus, as in (27), or end weight, as in (28) (cf. Quirk et al. 1972):

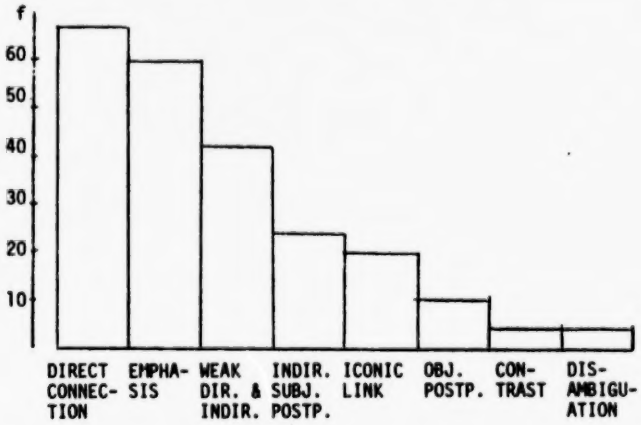
(27) In the place of lesser honour, because leg to leg with his more ordinary and less dramatic limb, was Ellen Mulliner. (WL p.148.)

(28) About that there were no two ways, thought this great keen-eyed kid from the outer darkness. (WK p. 109.)

The last diagram, Diagram 7, next page, presents frequencies of topicalization triggers. In other words, it gives some indications as to when a topicalization is possible and even likely. There seems to be a tendency to front a constituent if it provides a direct link to the preceding context, as in (29); if it requires a more emphatic position, as in (30); or if there are conspiracies involving multiple reasons for an unusual or more marked position, as in (31).

(29) Against that there was no appeal. (WL p. 87.)

Diagram 7. Frequencies of topicalization triggers (N=243).



(30) To be standoffish you must have dignity. (WL p. 95.)

(31) Outside the door through which they had just passed a man was kneeling, half in and half out of the moonlight. (WL p.49.)

Finally, if we wish to see whether the triggering mechanisms are different for different types of topicalized elements within the corpus, some hypotheses for further study can be drawn from Table 1. Here a cross-tabulation of initial constituent type and topic trigger has been converted into percentages. Direct objects were topicalized either when they provided a direct connection with a previous item, or for reasons of emphasis; the figures for all topicalized elements together were 27 and 26 per cent respectively, whereas they were 41 and 35 per cent for direct objects alone. Topicalized complements and topicalized valency adverbials tended to have either direct or emphatic connections, whereas other continuous adverbials and verbs were placed initially mainly for emphasis. The combined direct and indirect reasons most frequently affected adverbials of setting and other continuous adverbials, the figures being 25 and 27 per cent compared with the average 17 per cent for all topics. The topicalizations that were triggered off by iconic reasons usually fronted a direct or indirect object.

As these examples show, a great deal of suggestive information can be gathered from such cross-tabulations. But the present pilot

Table 1. Topic trigger for different types of topicalized constituents (in percentages) (N=243).

TOPIC TRIGGER	DIRECT CONNECTION	EMPHASIS	CONTR. OR CORR. CONN.	INDIR. REASONS SUBJ.POSTP.	INDIR. REASONS OBJ. POSTP.	DIR. & INDIR. REASONS	ICONIC REASONS	DISAMBIGU-ATION
TYPE OF CONSTITUENT								
DIRECT OBJECT	41	35	0	0	0	6	18	0
INDIRECT OBJECT	30	30	10	0	10	0	20	0
COMPLEMENT	53	16	5	26	0	0	0	0
VALENCY ADVERBIAL	44	26	0	17	4	4	4	0
SETTING ADVERBIAL	23	12	2	14	5	25	16	2
OTHER CONT. ADVERB.	17	35	1	4	2	27	10	3
SEPAR. ELEMENT	32	26	0	5	21	10	5	0
VERB	17	33	8	17	0	17	8	0
SUBJ. OF DEP. CLAUSE				100				
ADJ. CLAUSE	33	67						
ADV. CLAUSE	100							
ALL TOPICS	27	26	2	10	5	17	10	2

materials were too restricted and the coding criteria still too undeveloped to allow firm generalizations. Therefore the present report should be read as a pilot study with mainly methodological relevance rather than as a definite grammatical treatise.

6

The other way of testing contextual adjustment and textual fit is by acceptability experiments. In principle one could try two methods: present the same sentence in different contexts, or a set of different sentences in the same context. In either type of experiment, a group of suitably selected subjects could then be asked to pronounce on contextual

acceptabilities (by asking them for instance which sentence they prefer in the context given, or in which of the contexts a given sentence seems most fluent or natural). For reasons of space, only one simple pilot experiment will be discussed here; it involved placing different sentences into the same textual environment. The text was

> Like Lee, Stonewall Jackson had served gallantly in the Mexican War. He had devoted himself to the theoretical study of the military art.
>
> (1) At the Virginia Military Institute, Jackson was at this time a professor.
> (2) A professor at the Virginia Military Institute Jackson was at this time.
> (3) Jackson was at this time a professor at the Virginia Military Institute.
> (4) At this time Jackson was a professor at the Virginia Military Institute.
> (5) A professor Jackson was at this time at the Virginia Military Institute.
>
> He came of Ulster stock, settled in Virginia.

Twelve native-speaker lecturers in English from six different universities in Finland were asked to rate the contextual acceptabilities of the test sentences, first ranking them in order of preference, and then placing them on a three-point scale of acceptability. Half of the subjects regarded (4) and half (3) as their first choice, and the second choice was also evenly divided between (4) and (3). The third choice was (1), unanimously; the worst sentences were (2) and (5). The acceptability ratings went like this:

Sentence	All right	Doubtful	Unacceptable
1	2	7	3
2	0	1	11
3	9	3	0
4	10	2	0
5	0	0	12

This experiment is suggestive and can teach the investigator a lesson. The subjects were asked to comment and justify their judgments and to try to write contexts into which the ill-fitting sentences could be placed more snugly than into the context actually given. The comments showed that some of the subjects (professional teachers of English to foreigners) may well have worried about grammatical correctness and aired their own linguistic prejudices. The context given was also said to have been too short

to provide a basis for firm judgments. When asked to provide plausible contexts improving the acceptability of the awkward sentences, the subjects ventured solutions such as:

> Like Lee, Stonewall Jackson had served gallantly in the Mexican War. He came of Ulster stock, settled in Virginia. At the Virginia Military Institute, Jackson was at this time a professor.

This justifies the topicalization by linking Virginia Military Institute to Virginia in the previous sentence and to war in the preceding one. Some of the recontextualizations were fairly elaborate and show that the subjects were willing to adjust to the spirit of the game:

> Jackson was meanwhile pursuing his Jekyll-and-Hyde existence with unremitting energy. At the Virginia Military Institute, Jackson was at this time a professor. In Baltimore, Md., he was a dealer in antiques, while in Philadelphia, Pa., he was masquerading as a gynecologist.

> In the course of his life Jackson ran through a number of careers. A bootblack boy in his childhood, a bank messenger in Baltimore in his teens, a professional soldier as a young man and subsequently a yeoman farmer in Tennessee. A professor at the Virginia Military Institute Jackson was at this time. Later he played the roles of politician and judge, before retiring as a decadent aristocrat.

These solutions try to justify the topicalization through iconic parallellism with neighbouring sentences. Several informants also ventured comments such as:

> On sentence 1: "He made a slow start at Yale, but then made rapid progress: at the VMI Jackson was ... a professor." In this case, though, the following sentence scarcely fits. And in none of the sentences is it clear what 'at this time' refers to - surely not the Mexican War? More context needed.

The subjects were also asked for comments on the test itself, without being told precisely what the purpose of the exercise was. Several of the testees made critical suggestions: there was not enough context, the style was not controlled (several informants volunteered that some thematic variants looked like Irish or Welsh English), many sentence variants were so flagrantly unacceptable as to be hardly worth including, and so forth. "My most serious worry about this kind of test," one subject wrote,

> is whether suitable native speakers can be relied upon to evaluate some of these examples, and especially to suggest contexts ... It may be that naive native speakers, and especially those of other language backgrounds (e.g., immigrants of first or possible second generation) would be somewhat more lenient than a 'good' prose writer in what they would allow and rather at a loss when trying to make up contexts for examples - an 'unnatural' text-creation process.

There is, then, no ideal informant: the naive are too naive, the sophisticated are too sophisticated. This conflict between norms and usage has also been anticipated by Joseph Grimes, who wrote in his book The Thread of Discourse as follows, having included structures such as Saw the play did John, Saw the play did John? and The play did John see? among the normal paradigms of English:

> Before he reacts to the examples ... as bad English, as many do the first time they encounter them, the reader should observe his own thematizing behaviour for a day or so. Our grammatical tradition is heavily biased toward regarding unmarked thematization as well behaved and proper, and marked thematization as aberrant. Not so; it is part of the language, used constantly, related systematically to the rest. In fact, many of the phenomena written off in grammars as free word order are in contrast with each other thematically. (Grimes 1976: 330-331.)

However crude such acceptability experiments may be, they do show that groups of suitably selected informants can rate acceptabilities of thematic sentence variants with a reasonable degree of agreement, and that they can be coaxed to produce contexts which improve the acceptabilities of specific, otherwise doubtful thematic variants. Iconic parallellisms and the manipulation of old and new information were the two devices resorted to in the experiment referred to here.

7

In more general terms, the gist of our argument is as follows:

Certain word-order patterns of English are more basic and less marked than others. The more strongly marked a pattern seems to be in isolation, the stronger must be the contextual forces motivating its use, if it is to strike us as natural in a text. One basic principle of text strategy is based on distinctions between old and new information: usually a text proceeds from old information to new, unless it is to

call especial attention to the reversal of the information flow through clefts, marked focus, emphasis, or tricks of style. When arranging his sentences to obey the strategy and to organize old and new information, a speaker or writer has at his disposal lexical devices, syntactic devices involving thematization, and topicalizations. Lexical devices involving converses and subject assignment are presumably the least conspicuous and obtrusive. We do not always stop to think that the difference between sentences such as those in (4) and (5) (p.51) is due to thematization at all. Other devices may strike us as being more strongly marked, and in fact they can be said to form a scale of markedness. Topicalizations of certain adverbials (adverbials of setting), for instance, are relatively weakly marked. Topicalizations of objects need stronger motivation, and topicalizations of parts of verb phrases still stronger motives. In addition to the strategic principles concerning the distribution of old and new information, there are other forces motivating thematic and topical arrangements. One is end weight: weighty and long constituents often go last, perhaps because they owe their weight to the new information. Iconic motives are the strongest of all, where they exist. Thus requirements of syntactic parallellism and metre can override all other considerations in such styles as are devoted to their pursuit; the freedom to use strongly marked thematic patterns is an important aspect of poetic licence.

Thus, if the motives are strong enough, they may justify the textual fit even of strongly marked sentence patterns. And this can be linked to what was said above in Section 3 about paradigmatic and syntagmatic markedness. What is strongly marked at the clause level, and thus strongly marked when the sentence occurs in isolation, can be unmarked at the textual level. There are situations and styles that are willing to tolerate clause and sentence patterns which would look weird in isolation, even to the point of deviance and beyond, as long as they are not too strongly marked in the syntagmatic environment of the text. But to ensure syntagmatic acceptability and textual fit, the thematic pattern must be strongly motivated and consonant with the text strategy.

To what extent these principles are universal is an open question. Certainly different languages give very different degrees of importance to the informational organization of the sentence. Li and Thompson (1975) have drawn attention to the typological significance of information dynamics, exemplifying their tenets with references to lan-

guages distant in geographical distribution and provenance. But it might be worth adding that even such relatively closely related languages as English and Swedish are thematically different, English being thematically more rigid and Swedish thematically more flexible. Such observations open up new areas of application. In foreign-language teaching, differences in the handling of information flow in the source and target languages may deserve more explicit attention, particularly in instances where these languages make use of different mechnisms. And in translation, thematic equivalence -- that is, a similarity in text strategy and information flow -- is an important criterion of translational equivalence and adequacy. Translators too are among those who ought to learn more about contrastive thematics, about the relative markedness of thematically different sentence patterns, and about the forces improving or impairing textual fit. To what extent these matters can be stated as rules, and to what extent they remain mere trends and tendencies, is a problem we must leave for another occasion.

BIBLIOGRAPHY

Enkvist, Nils Erik. 1975. Tekstilingvistiikan peruskäsitteitä. Helsinki: Gaudeamus.

--- 1976. "Notes on Valency, Semantic Scope, and Thematic Perspective as Parameters of Adverbial Placement in English," in N.E. Enkvist & Viljo Kohonen, eds., Reports on Text Linguistics: Approaches to Word Order. Publications of the Research Institute of the Abo Akademi Foundation, 8, Abo, pp. 51-74.

--- 1977. "Contextual Acceptability and Error Evaluation," in Rolf Palmberg & Håkan Ringbom, eds., Papers from the Conference on Contrastive Linguistics and Error Analysis, Stockholm & Abo, 7-8 February, 1977. Publications of the Research Institute of the Abo Akademi Foundation, 19, Abo, pp. 1-26.

Greenbaum, Sidney. 1976. "Syntactic Frequency and Acceptability," Lingua 40:2/3, 99-113.

--- 1977. "The Linguist as Experimenter," in F.R. Eckman, ed., Current Themes in Linguistics. New York: Hemisphere Publishing Corporation, pp. 125-144.

--- & Randolph Quirk. 1970. Elicitation Experiments in English. London: Longman.

Grimes, Joseph E. 1975. The Thread of Discourse. (= Janua Linguarum, Series Minor, 207.) The Hague & Paris: Mouton.

Halliday, M.A.K. 1967. "Notes on Transitivity and Theme in English.II," Journal of Linguistics 3, 177-274.

Kohonen, Viljo & Jussi Salmela. 1978. "Aineiston valinnan ja automaattisen tietojenkäsittelyn ongelmia kielitieteellisessä tutkimuksessa", this vol.

Li, Charles N. & Sandra A. Thompson. 1975. "Subject and Topic: A New Typology," in Charles N. Li, ed., Subject and Topic. New York etc.: Academic Press, Inc., pp. 457-490.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1972. A Grammar of Contemporary English. London: Longman.

Sköld, Tryggve. 1970. The Object in Pre-Predicative Position in Swedish (= Acta Universitatis Upsaliensis. Acta Societatis Linguisticae Upsaliensis, Nova Series 2:3), Uppsala.

BEST COPY AVAILABLE